ICARDA Agro-Climate Tool Technical Description

Steven A. Mauget
USDA-ARS Plant Stress and Water Conservation Laboratory
Lubbock, Texas

Eddy De Pauw
International Center for Agricultural Research in the Dry Areas
Aleppo, Syria

Contents

1. Introduction

The ICARDA Agro-Climate Tool (hereafter, 'the application') is a Visual Basic (6) program that can be run on Windows 98, 2000, and XP operating systems. It should be installed on a PC with a Pentium III or better microprocessor and at least 230 Mbytes of available hard disk space. Monitor screen resolution should be at least 1024 X 768 pixels but no more than 1920 X 1440 pixels. Instructions for the application's use can be found by left single-clicking on 'Instructions' on the application's upper left corner.

The application's primary daily variables (daily minimum temperature, daily maximum temperature, precipitation) were generated by modified GEM6 (Hanson et al., 1994) weather generator code. For reference to primary variable generation, see the flow chart in Fig. 2. Secondary variables (daily dew point temperature, short-wave surface radiation, net outgoing long-wave radiation and reference grass evapotranspiration) were derived from primary variables using algorithms drawn from the FAO's 'Guidelines for Computing Crop Water Requirements' (Allen et al., 1998), hereafter referred to as 'FAO-56'. Crop evapotranspiration values were then derived from the reference grass ET values using the FAO-56 single crop coefficient method. For reference to secondary variable generation, see the flow chart in Fig. 6.

3. Data

The application's climate statistics are derived from two data sets that provide daily records of minimum and maximum temperature and precipitation. The main data source is the Global Daily Summary Data (GLDS) set (National Climatic Data Center, 1994), which provides data for ICARDA growing regions from 590 meteorological station locations. The period of record for this data set is October 1977 to December 1991. The secondary data source is the Global Daily Climatology Network data (National Climatic Data Center, 2002), which provides records of primary daily variables at 59 locations (Fig. 1). Data from GDCN stations is of varying duration, but in some cases begins in the early 20th century. However, because the application's operation involves the averaging of weather generator parameters from different stations to derive parameters for locations between stations, those parameters must be derived from data over a uniformly defined period. As a result, the decision was made to limit the calculation of statistics from GDCN data to the GLDS data period, i.e., 1977-1991. Future versions of the application will attempt to expand data coverage to a longer data period.

3. Statistics Calculation

3.1 Sampling Requirements and Uncertainty

The lack of long-term daily station data, and the sparse nature of the data that was available over the ICARDA mandate region, was a primary limiting factor in the calculation of the climate statistics the application is based on. One of the leading development challenges involved deriving climatologically representative streams of stochastic weather variability from relatively short, and sometimes fragmentary, weather records. This challenge was met mainly through modifications to the original GEM6 code of Hanson et al., (1994) and by imposing minimum data sampling requirements. An addi-

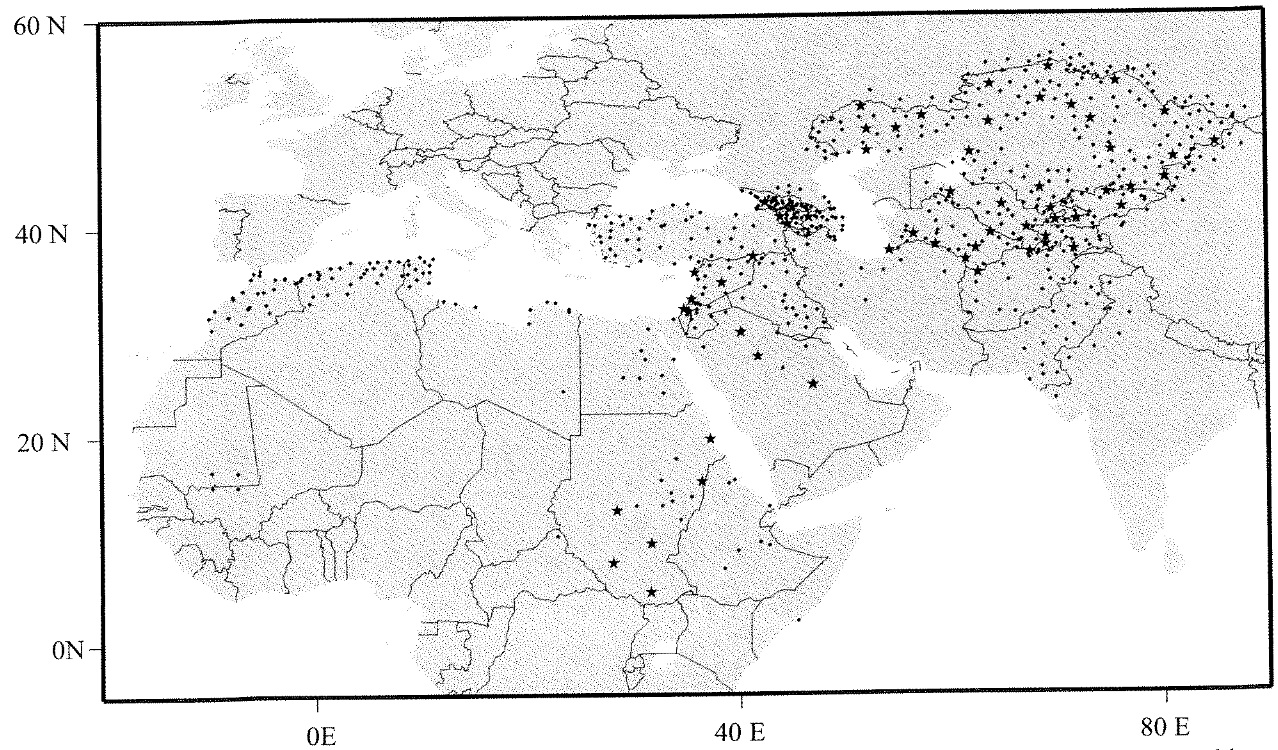


Figure 1. Locations of the 590 GLDS (dots) and 59 GDCN (star) meteorological stations used here.

tional strategy for addressing this problem requires the end-user to use the application in a way that acknowledges the possibility of the resulting sampling uncertainties.

Normally, GEM6 code calculates weather statistics over 24 bi-weekly periods of the year, but the limited daily data availability over ICARDA agriculture areas made bi-weekly averaging impractical. As a result, the GEM6 weather generator parameters calculated here were derived from monthly statistics of precipitation, and minimum and maximum temperature. A complete daily data record during October 1977–December 1991 would result in approximately 14*30 = 420 daily weather measurements contributing to each month's precipitation and temperature statistics. However, in ICARDA agricultural areas outside of the former Soviet Union data gaps were frequent. To provide adequate station coverage over those areas while also calculating reasonably representative monthly statistics, a minimum sampling threshold of 60 days for monthly statistics was imposed. Thus for example, the mean daily maximum temperature for January at a station location might be based on as few as 60 daily temperature measurements. Deriving an average from such a relatively small number of measurements can lead to the following biases:

1. <u>Sampling error in the resulting statistic</u>. The magnitude of this sampling error is proportional to N^{-2}, where N is the number of measurements (Mendenhall et al., 1990). A sample mean calculated from N=60 measurements can lead to an error

(i.e., the difference between the true, population mean and the mean calculated from a 60 day sample) of the order of .13 * σ, where σ is the standard deviation of the population distribution. However, this error can be as large as .26 * σ. Errors for monthly rainfall transition probabilities (i.e., p00, p10 in Section 3.2), and daily temperature cross correlation values used in the GEM6 multivariate temperature generation scheme (i.e. Eq.s 12-15 in Hanson et al., 1994) may be of similar magnitude (See Appendix A).

2. Statistics representative of a limited subset of years during 1977-1991. Limiting the calculation of monthly statistics to 60 or more days makes it possible that those days may be, in the worst case, from two consecutive years, e.g. Jan. 1979 and Jan. 1980. Under those circumstances, the resulting statistics would not be representative of 1977-1991, and could differ considerably from nearby stations that reported more January data during the 1977-1991 period.

The method suggested here for identifying these possible errors is based on the assumption that they will most likely not be consistently evident at neighboring station locations. In operating the application the user selects a location by left-clicking the location on the large map. The nearby stations whose Fourier parameters will be used to estimate the location's GEM6 parameter set will then flash in sequence. ***In practice, the user should always compare the application's results for a location with the corresponding results for each of those nearby stations. The user should also compare each of the nearby station's results with one another.*** For example, the annual cycles for the probability of heat and cold stress and the probability of exceedance curves for precipitation and growing degree days might be compared. If these comparisons show one station's results to differ clearly from the remaining stations, that station's GEM6 parameters may have been derived from biased statistics. In that case the user should consider using results derived only from one of the remaining nearby stations.

3.2 Calculation of Monthly Statistics

The limitations of daily data over ICARDA agricultural regions also influenced the choice of statistics that were calculated from the data. The original GEM6 code calculates two sets of temperature statistics: the mean and standard deviation of maximum and minimum temperature during dry days, and the mean and standard deviation of maximum and minimum temperature during days with rain. However, dividing the temperature data this way would have caused many stations to fail the minimum monthly sampling requirement described above. As a result, temperature statistics here were calculated over all days, both wet and dry. For each station, temperature variation throughout the year was described through four sets of statistics:

1. Mean maximum daily temperature (TMXM) for month i, i = Jan. Feb., ...Dec.
2. Standard deviation of maximum daily temperature (TMXS) for month i, i = Jan. Feb.,...Dec.
3. Mean minimum daily temperature (TMNM) for month i, i = Jan. Feb., ...Dec.
4. Standard deviation of minimum daily temperature (TMNS) for month i, i = Jan. Feb.,...Dec.

GEM6 uses multivariate regressive and autoregressive relationships to derive daily anomalies of maximum and minimum temperature (tmax, tmin) and short-wave surface radiation (srad) based on the current and previous day's anomaly values (Eq. 12-15 Hanson et al. (1994)). The 3 X 3 'A' and 'B' matrices defining these regressive relationships are derived from cross correlation values calculated between tmin, tmax, and srad at 0 and 1 days lag. But because daily srad values are not available over the ICARDA mandate region, temperature generation here is based on 2 X 2 matrices that are derived from cross correlation values calculated only between daily tmin and tmax values. Although GEM6 calculates 'A' and 'B' matrices for each month, this application calculates only one 'A' and 'B' matrix per station, which are in turn derived from annual averages of tmin and tmax correlation and cross correlation values. These matrices are formed from 8 values:

5. A(1,1), A(1,2), A(2,1), A(2,2)
6. B(1,1), B(1,2), B(2,1), B(2,2)

The probability that a day will be rainy in the GEM6 generation scheme depends on two sets of monthly statistics:

7. The probability that a dry day will be followed by a dry day (p00) during month i, i= Jan., Feb., ...Dec.
8. The probability that a wet day will be followed by a dry day (p10) during month i, i= Jan., Feb., ...Dec.

GEM6 code normally assigns the amount of rain that falls on a wet day using a mixed exponential distribution. However, here it was found that the three-parameter mixed exponential distribution did not perform noticeably better than a simple one-parameter exponential distribution. As the parameter for an exponential distribution is the expectation of the distribution's variable, (page 167-68, Mendenhall (1990)), the exponential parameter used here is the average of the daily rainfall totals, calculated for each month of the year.

9. The mean of daily rainfall totals (XMU) during month i, i = Jan., Feb.,...Dec.

4. GEM6 Fourier Parameter Calculation and Storage.

The annual cycles of monthly statistics described in Section 3 above (i.e., statistics 1-4 and 7-9) are interpolated to daily variability in the application by solving for the first three annual Fourier harmonics, and then using those harmonics to reconstruct a smoothed version of the monthly annual cycle through an inverse transform. The results of the Fourier transform are stored here in the Access Database 'Data/PAR_GEN.mdb' in the table 'par_gen'. That table's contents provides the primary inputs for the application's daily weather generation scheme. The table consists of 649 rows, one for each meteorological station. Each row consists of 62 columns. Columns 1-5 identify the station and its geographic coordinates:

Column 1:	Station Index ('stnindx' = 1-649)
Column 2:	11 character Station Identifier ('stnid')
Column 3:	Station Longitude ('stnlon') with degrees E. > 0, degrees W. < 0.
Column 4:	Station Latitude ('stnlat').
Column 5:	Station Elevation ('stnelev') in meters.

The results of the Fourier transforms of Section 3's 1-4 and 7-9 statistics are stored in columns 6-54. The Fourier transform for each annual cycle of monthly statistics produces seven real numbers: the mean of the annual cycle, the amplitudes of the first three annual harmonics and the phase angles of the first three annual harmonics. Columns 6-12 store the results for the 'p00' statistic:

Column 6:	Annual Mean of p00 ('p00mean').
Column 7:	First Harmonic Amplitude for p00 ('p00amp1').
Column 8:	Second Harmonic Amplitude for p00 ('p00amp2').
Column 9:	Third Harmonic Amplitude for p00 ('p00amp3').
Column 10:	First Harmonic Phase Angle for p00 ('p00pan1').
Column 11:	Second Harmonic Phase Angle for p00 ('p00pan2').
Column 12:	Third Harmonic Phase Angle for p00 ('p00pan3').

Columns 13-54 store the Fourier parameters for the remaining 6 annual cycles of monthly statistics.

Columns 13-19:	Fourier Parameters for p10 ('p10mean'–'p10pan3')
Columns 20-26:	Fourier Parameters for XMU ('XMUmean'–'XMUpan3')
Columns 27-33:	Fourier Parameters for TMNM ('TMNMmean'–'TMNMpan3')
Columns 34-40:	Fourier Parameters for TMNS ('TMNSmean'–'TMNSpan3')
Columns 41-47:	Fourier Parameters for TMXM ('TMXMmean'–'TMXMpan3')
Columns 48-54:	Fourier Parameters for TMXS ('TMXSmean'–'TMXSpan3')

Columns 55-62 store the individual elements of the 'A' and 'B' matrices.

Column 55:	A(1,1) ('am11').
Column 56:	A(1,2) ('am12')..
Column 57:	A(2,1) ('am21')..
Column 58:	A(2,2) ('am22')..
Column 59:	B(1,1) ('bm11')..
Column 60:	B(1,2) ('bm11')..
Column 62:	B(2,1) ('bm11')..
Column 62:	B(2,2) ('bm11')..

5. GEM6 generation of primary synthetic variables (Figure 2).

5.1 Spatial Interpolation of GEM6 Parameters Between Stations

When the user selects a location by left-clicking on a pale yellow area of the large map, a number of nearby stations will flash in sequence. The VB6 code then calculates that location's GEM6 parameters as an inverse-distance weighted average of the Fourier parameter sets and the 'A' and 'B' matrix elements of those neighboring stations. Sets of neighboring stations for ICARDA agricultural areas - the yellow shaded areas in the large map - are defined in the table 'near_neighbor' in the Access Database 'Data/PAR_GEN.mdb'. That table divides the yellow area into 792 1° longitude by 1° latitude grid areas. The neighboring stations for a 1° by 1° grid area are the stations that lie within a 3° longitude by 3° latitude grid that surrounds that central 1° by 1° grid. The 'near_neighbor' table contains a row for each 1° by 1° grid, each of which consists of 26 columns.

Column 1:	The grid index of the 1° by 1° grid ('grdindx' = 1-792).
Column 2:	The number of stations in the surrounding 3° by 3° grid ('nstn').
Column 3:	The longitude of the center of the 1° by 1° grid ('grdlon').
Column 4:	The latitude of the center of the 1° by 1° grid ('grdlat').
Column 5:	An integer longitude index for the 1° by 1° grid ('ilon').
Column 6:	An integer latitude index for the 1° by 1° grid ('ilon').
Columns 7-26:	The station indices (i.e., the identifying station integer values in column 1 of the table 'par_gen') for each of the stations in the surrounding 3° by 3° grid ('stnindx1 – stnindx20')

Once the VB6 code determines which 1° by 1° grid contains the selected location, the Fourier parameter sets for the grid's neighboring stations listed in columns 7-26 are then retrieved from the PAR_GEN database 'par_gen' table. The distances between the selected location and those stations are then calculated, and the stations are then sorted according to their distance from the selected location. If the nearest station is within 20 kilometers, then that station's parameters are assigned to the location. Otherwise, the location's parameters are calculated as a distance weighted average of the nearest neighboring stations using an expanding radius search method.

- If the two nearest stations are within 40 km, then those stations parameters are averaged,
- Else, if the three nearest stations are within 60 km, then those stations parameters are averaged,
- Else, if the four nearest stations are within 80 km, then those stations parameters are averaged,
- Else, if the five nearest stations are within 100 km, then those stations parameters are averaged,
- If there are more than eight nearby stations, the location's parameters are averaged from the nearest eight.
- If none of these conditions are met, the location's parameters are averaged from all of the stations in the surrounding 3° by 3° grid.

In some areas where station coverage is sparse (e.g., Sudan and Ethiopia) the last condition could cause a location's parameters to, in the worst case, be averaged from the parameters of stations ~ 200 km away.

5.2 Maximum and Minimum Daily Temperature Generation

To account for the effects of elevation on a selected location's interpolated temperature variation, the mean temperature Fourier parameters (i.e., 'TMNMmean' and 'TMXMmean') of the neighboring stations are adjusted to sea-level before inverse-distance averaging. That is,

$$\text{TMNMmean} = \text{TMNMmean} - \text{Station Elevation} * \text{lapse rate}, \qquad (1)$$

$$\text{TMXMmean} = \text{TMXMmean} - \text{Station Elevation} * \text{lapse rate},$$

where a mean wet adiabatic atmospheric lapse rate of – 6.5° C/Km is assumed. The adjusted mean temperature parameters, and all the remaining amplitude and phase angle Fourier parameters for all the stations contributing to a location average are then averaged using an inverse distance^2 averaging scheme. Thus TMXMmean at the selected location is calculated as,

$$\text{TMXMmean}_{\text{Loc}} = \frac{\sum_i \frac{1}{d_i^2} * \text{TMXMmean}_i}{\sum_i \frac{1}{d_i^2}} \qquad (2)$$

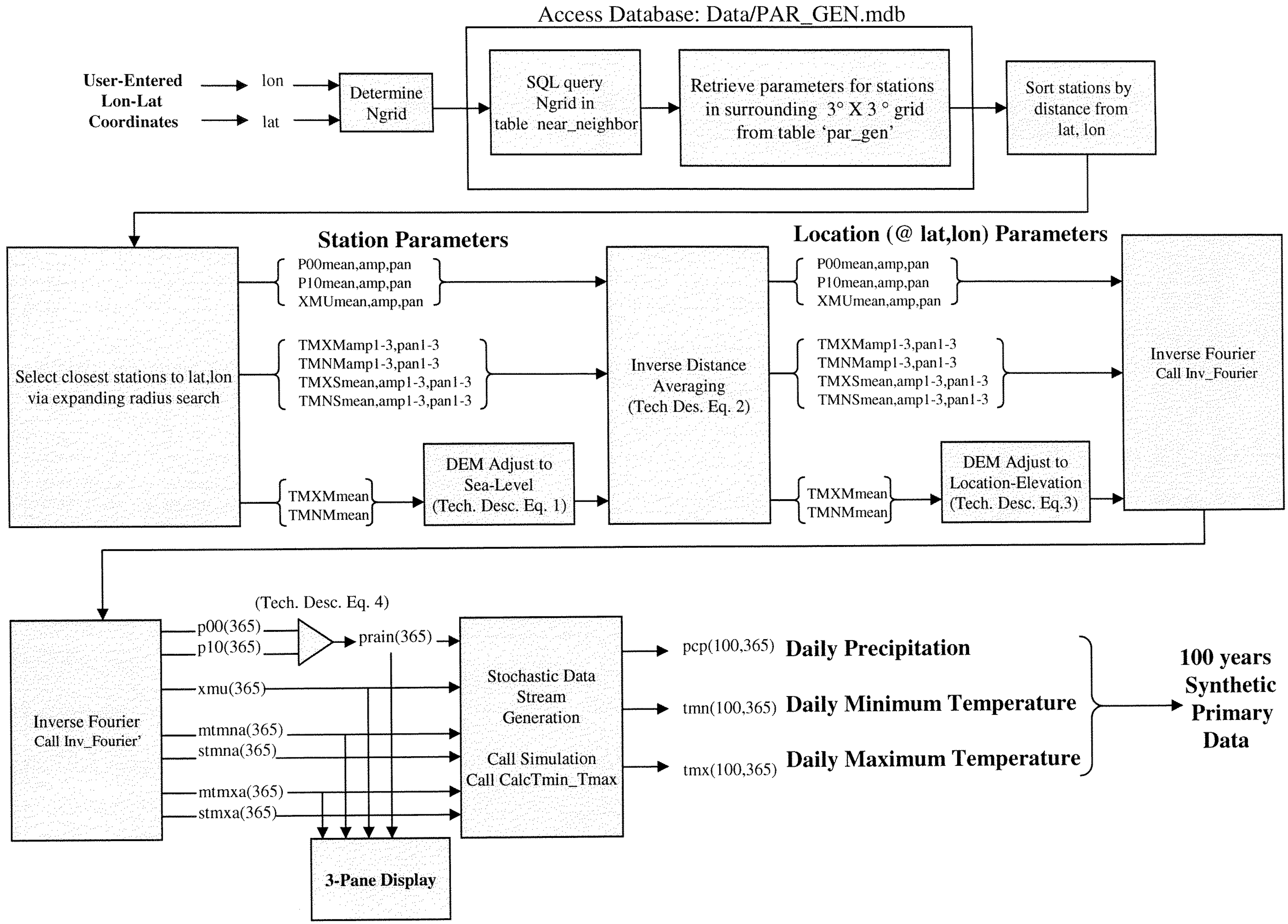
Where d_i is the distance between station i and the selected location, and TMXMmean_i is the TMXMmean value for station i. After all of the location's Fourier parameters have been estimated in this way, the location's mean maximum and minimum temperature parameters at sea level are then adjusted – in most cases decreased, as most stations are above sea-level – to the location's elevation as defined by the GTOPO30 digital elevation model.

$$\text{TMNMmean}_{\text{Loc}} = \text{TMNMmean}_{\text{Loc}} + \text{GTOPO30 Location Elevation} * \text{lapse rate}, \quad (3)$$

$$\text{TMXMmean}_{\text{Loc}} = \text{TMXMmean}_{\text{Loc}} + \text{GTOPO30 Location Elevation} * \text{lapse rate},$$

After the entire Fourier parameter set for the selected location has been spatially interpolated, the location's annual cycles of the mean and standard deviations of daily maximum temperature and of the mean and standard deviations of daily minimum temperature are then constructed through an inverse Fourier transform. The annual cycles of mean daily maximum and minimum temperature are used to depict the location's annual temperature variation in the top graph of the Three-Pane display on the application's left side. The four annual mean and standard deviation annual cycles, and the location's spatially interpolated 'A' and 'B' matrices, are then used to generate stochastic streams of daily maximum and minimum temperature in the subroutine 'CalcTmax_Tmin'. These streams are stored as 100 years of synthetic temperature variation in the arrays tmn(100,365) and tmx(100,365).

Fig. 2. ICARDA Agro Climate Tool Application Flow Chart: Primary Synthetic Data Generation



5.3 Daily Precipitation Generation

A selected location's precipitation Fourier parameters (i.e., the 'p00', 'p10' and 'XMU' mean, amplitude, and phase angle parameters) are estimated using the inverse distance weighting method of Eq. 2. The resulting parameters for the location are then inverse transformed into three annual cycles: the annual cycle of p00, the annual cycle of p10, and the annual cycle of the exponential rainfall distribution parameter (XMU). The annual cycle of the probability that rain falls on a given day of the year is derived from the p00 and p10 probabilities via Eq. 4 of Hanson et. al (1994).

$$P(\text{day n is wet}) = \frac{1 - p00(n)}{1 + p10(n) - p00(n)}, \quad n = 1, 365 \tag{4}$$

These probabilities for each day of the year are used to graph the location's annual cycle for daily rainfall probability in the middle graph of the Three-Pane display on the application's left side. Because the exponential rainfall distribution parameter is equal to the average rainfall amount on wet days, the XMU annual cycle is used to graph the average rainfall amount on wet days in the bottom graph of the Three-Pane display. The annual cycles of XMU and daily rainfall probability are both used to generate 100 years of synthetic precipitation data which is stored in the array pcp(100,365).

6. Comparison of real vs. generated primary variable statistics.

The application's purpose is to estimate the statistics of actual variation of daily maximum and minimum temperature and precipitation, and also secondary daily variables derived from those primary variables. But because a database containing multiple-year records of daily data from 649 weather stations would be too large to distribute, daily weather variation is generated here from a relatively small set of weather generator parameters derived from real data. As a result, the accuracy of the application's reported statistics is based on the equivalence of real data statistics with the statistics of the corresponding GEM6-generated data streams. That correspondence is checked here by comparing mean bi-weekly temperature and precipitation statistics (e.g., as in Fig. 3) derived from both real and generated data.

Figure 4a is a scatterplot of mean bi-weekly heating degree days (HDD) from the annual cycles of the 59 GDCN stations used here. Figure 4b is a counterpart figure for mean cooling degree days (CDD). An example of both of those annual cycles can be found in Fig. 3c. Figure 4's X-coordinates are the bi-weekly means of the real data, while the Y-coordinates are the means of the generated values for the same bi-weekly period. In both Figs. 4a and b there is relatively close agreement between real and generated bi-weekly degree-day means. The root-mean squared error for mean heating and cooling degree days is 10.48° C and 6.73° C respectively.

Less agreement is found in the comparisons of real vs. generated bi-weekly precipitation statistics. Figure 4c is a scatterplot of mean bi-weekly cumulative precipitation, while Fig. 4d is a scatterplot of the mean bi-weekly percentage of wet days. In Fig. 4c the higher degree of scatter for some larger values (> 20 mm) shows that the relative error (i.e., (real-generated)/real) for some bi-weekly averages can approach ~.40.

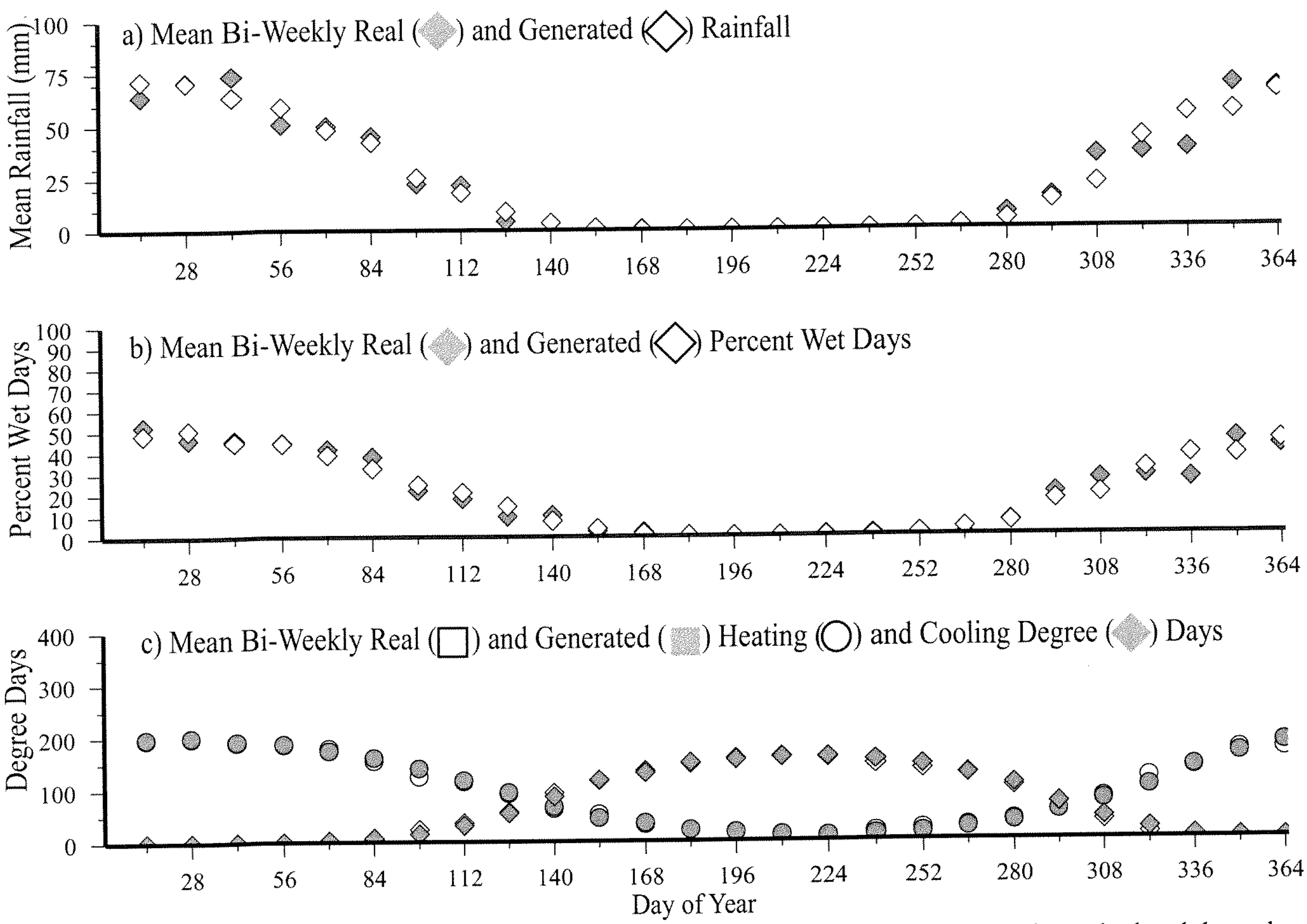


Figure 3. Bi-Weekly mean real and generated statistics for a representive GDCN station calculated throughout the year. a) Mean bi-weekly rainfall, b) Mean bi-weekly percent wet days, c) Mean bi-weekly heating and cooling degree days.

Somewhat less scatter is found in Fig. 4d, which shows that the GEM6 configuration used here is marginally better at reproducing rainfall frequency. However, Fig. 4c suggests problems with reproducing the amounts of daily rainfall during high rainfall periods. A scatterplot showing real vs. generated means derived from a GEM6 configuration using a mixed-exponential rainfall distribution (not shown) is qualitatively similar to Fig. 4c. The method suggested for checking the accuracy of the application's rainfall statistics is essentially that proposed in Section 3.1. That is, that the user should always cross-compare nearby station's results with one another to confirm consistency.

Figures 3 and 4 compare bi-weekly means, but the application presents information about the distribution of daily statistics accumulated over arbitrarily chosen periods. For example, cumulative growing degree days and heat stress and cold stress duration on the Temperature Tab and cumulative precipitation on the Precipitation Tab. While comparing real vs. generated primary data distributions over a representative number of CWANA stations is impractical here, Figure 5 makes such a comparison for Crosbyton, Texas

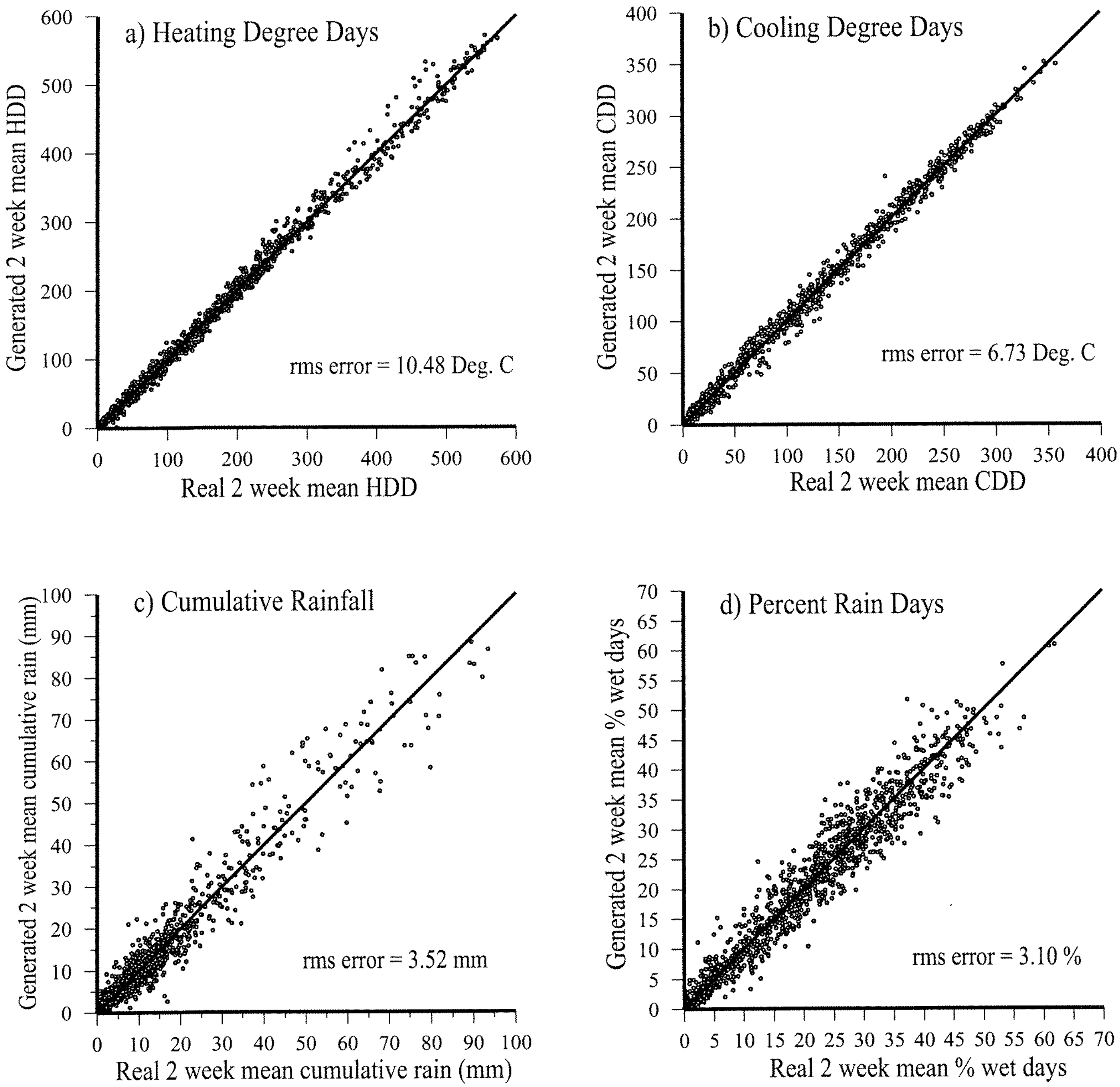


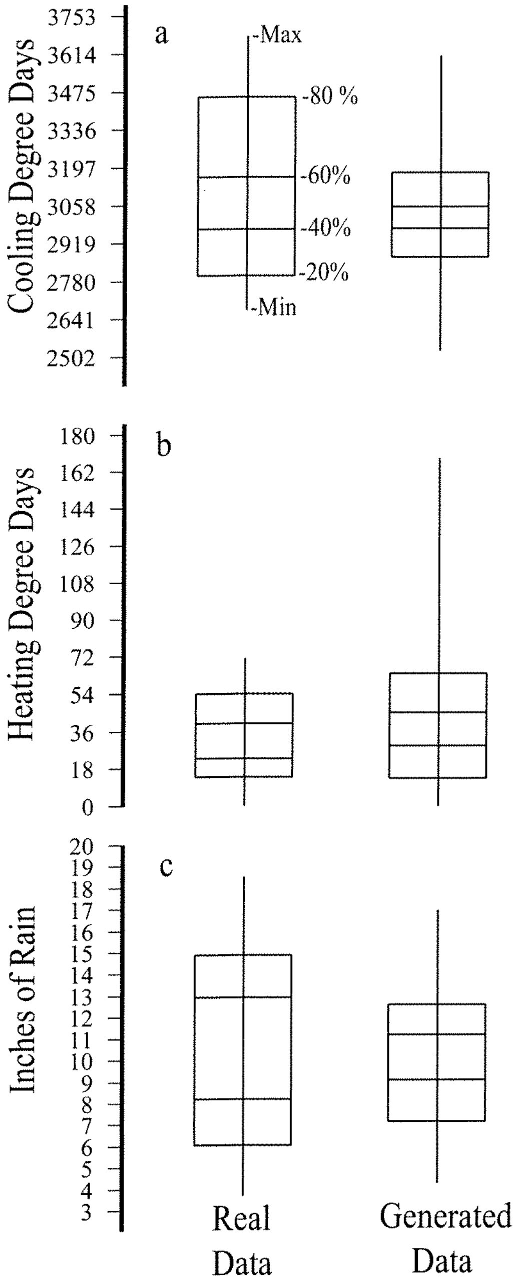
Figure 4. Scatterplots for bi-weekly statistics of real (x-axis) vs. generated (y-axis) daily temperature and precipitation data from the 59 Global Daily Climatology Network Stations used here. a) Mean bi-weekly heating degree days, b) Mean bi-weekly cooling degree days, c) Mean bi-weekly cumulative rainfall, and d), Mean bi-weekly percent wet days.

during the summer (May 15–Sept. 15) growing season. That station's daily temperature and precipitation data is relatively complete during 1971-2000, thus Fig. 5's real vs. generated comparisons reflect GEM6 characteristics rather than differences introduced by sampling uncertainty. Figures 5 a-c show bar-and-whisker plots of accumulated HDD, CDD, and rainfall for May 15-Sept. 15. Figures 5 d-g show probability distributions derived from real and generated data, and are analogous to the red, blue, green and gold bar-graphs on the application's Temperature and Precipitation Tabs.

In Figure 5a-c the whiskers mark the extremes of the data distribution, while the bar's horizontal lines mark the 80th, 60th, 40th, and 20th percentiles. In Figs. 5a and c the generated bars are narrower than their real counterparts, indicating that the distributions of generated cooling degree days and cumulative rainfall are narrower than that of the corresponding real data. This is consistent with the strictly random process by which weather generators produce synthetic data streams, and their related inability to reproduce interannual variation (Wilks, 1999). For example, weather generators such as GEM6 randomly introduce extreme wet days during the wet period of a station's annual rainfall cycle. However, for stations on the coast of California, extreme wet days tend to be clustered into winter periods marked by El Niño conditions. Other stations in areas subject to periodic drought years may experience abnormally long periods of high temperatures and no rain, which weather generators are unlikely to produce. As a result, the synthetic data distributions for those stations will show a lower probability of high cooling degree day totals and low rainfall totals. ***Weather generators are poorly suited to reproduce extreme variation in seasonal climate, thus generated probability distributions will tend to be narrower than their real counterparts.***

Figures 5d and e correspond to the red and blue bar graphs on the application's Temperature Tab with the heat and cold stress slider controls set to 90.0° F and 55.0° F. In Fig. 5d the GEM6 generated data shows a 17% probability of a 14 day or longer run of maximum daily temperatures greater than 90.0° F, while the incidence in the real data is 40%. Conversely, probabilities of runs of less than 9 days duration are higher in the generated data than the real data. This may be additional evidence of the problem described above, i.e., an inability of GEM6 to produce long runs of high daily maximum temperatures that are present in the real data. Figure 5e shows no occurrences of 9 day or longer runs of minimum temperatures less than 55.0° F in the real data, while the incidence in the synthetic data is ~15%. In this instance the generated data contains long runs of low minimum temperature conditions that are not present in the real data.

Figures 5f and g correspond to the green and gold bar graphs on the application's Precipitation Tab. Both of those figures show reasonable agreement between the daily rainfall amounts and the dry runs in both the generated and real data. However, Fig. 5's results apply to only one meteorological station and one period of the year. ***Generally, because of sampling uncertainty and the inherent shortcomings of weather generators, the application's probability of exceedance curves for crop evapotranspiration, precipitation, and growing degree days, and the bar graphs for temperature and precipitation, should be considered as approximate estimates.***



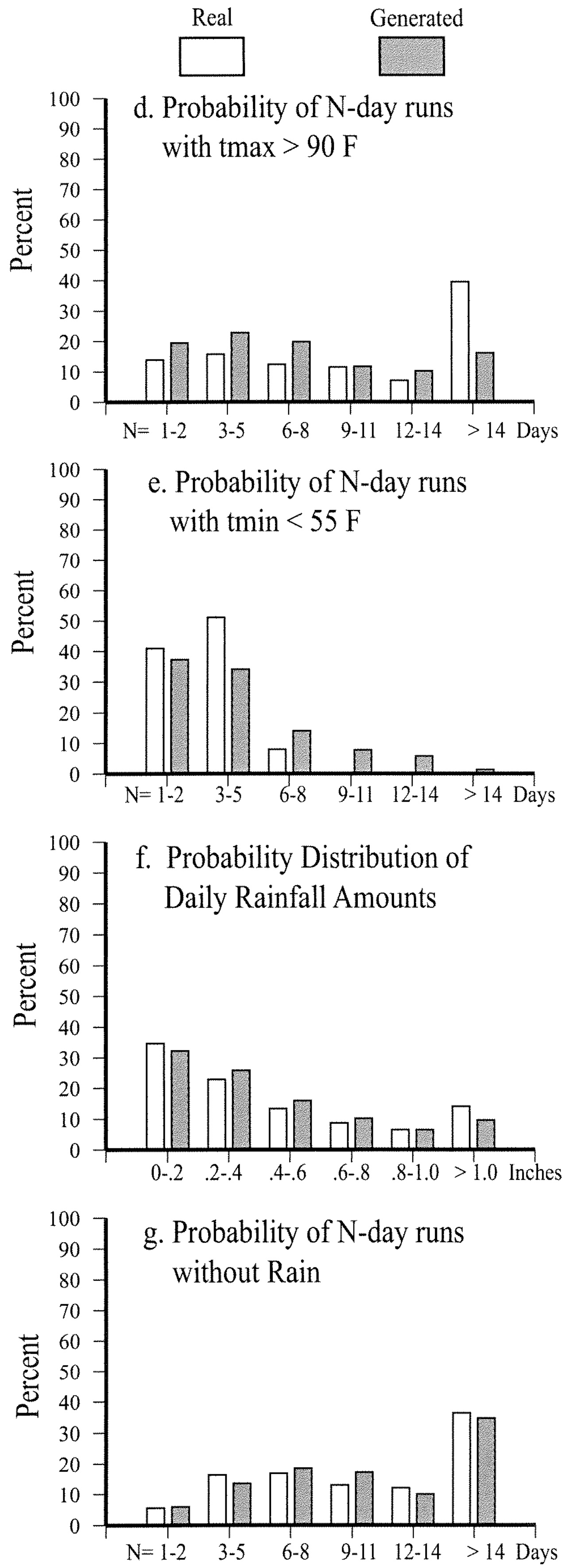


Figure 5. a-c) Box and whisker plots for distributions of cumulative cooling degree days, heating degree days and rainfall for May15 - Sept. 15 at Crosbyton, Texas. Distributions for actual daily data is on the left, for generated data on the right. d-g) Probability distributions for warm runs (d), cool runs (e), dry runs (g), and daily rainfall totals (f). White bars are real data distributions, gray bars are generated data distributions.

7. Secondary variable generation (Figure 6).

Secondary daily meteorological variables were derived here using parameterization relationships drawn from FAO-56 (Allen et al., 1998).

7.1 Dew Point Temperature

Daily dew point temperatures were estimated from daily minimum temperatures using the following parameterization scheme (FAO-56, Eq. 6-6):

- $T_{dew} = T_{min} - 2°C$ for locations in arid areas,
- $T_{dew} = T_{min}$ elsewhere.

Arid locations were defined here as stations with a mean annual temperature of 18°C or greater, and a mean annual number of wet days of 55 or less.

7.2 Shortwave radiation at the surface

Daily integrated shortwave surface radiation (R_S) was estimated using the Hargreaves radiation formula (FAO-56 Eq. 50):

$$\mathbf{R_S} = \mathbf{k_{Rs}} \sqrt{\mathbf{T_{max}} - \mathbf{T_{min}}}\ \mathbf{R_a} \tag{5}$$

Where,

- k_{Rs} is an adjustment coefficient, assigned here as 0.175 $°C^{-0.5}$,
- T_{min} is daily minimum temperature (in Celsius),
- T_{max} is daily maximum temperature (in Celsius), and,
- R_a is the daily integrated shortwave radiation at the top of the atmosphere in units of Joules * 10^6 / (met.^{2} * day) (FAO-56 Eq. 21).

7.3 Vapor Pressure and Saturation Vapor Pressure.

Given daily minimum, maximum, and dew point temperatures, the vapor pressure and saturation vapor pressure are solved for using the Clausius-Clapeyron equation ($e^o(T)$: FAO-56 Eq. 11)

$$\begin{aligned} &\text{Actual vapor pressure} = e_a = e^o(T_{dew}), \\ &\text{Saturation vapor pressure} = e_s = 0.5 * (e^o(T_{max}) + e^o(T_{min})) \end{aligned} \tag{6}$$

7.4 Net upwelling outgoing long-wave radiation (OLR) at the surface.

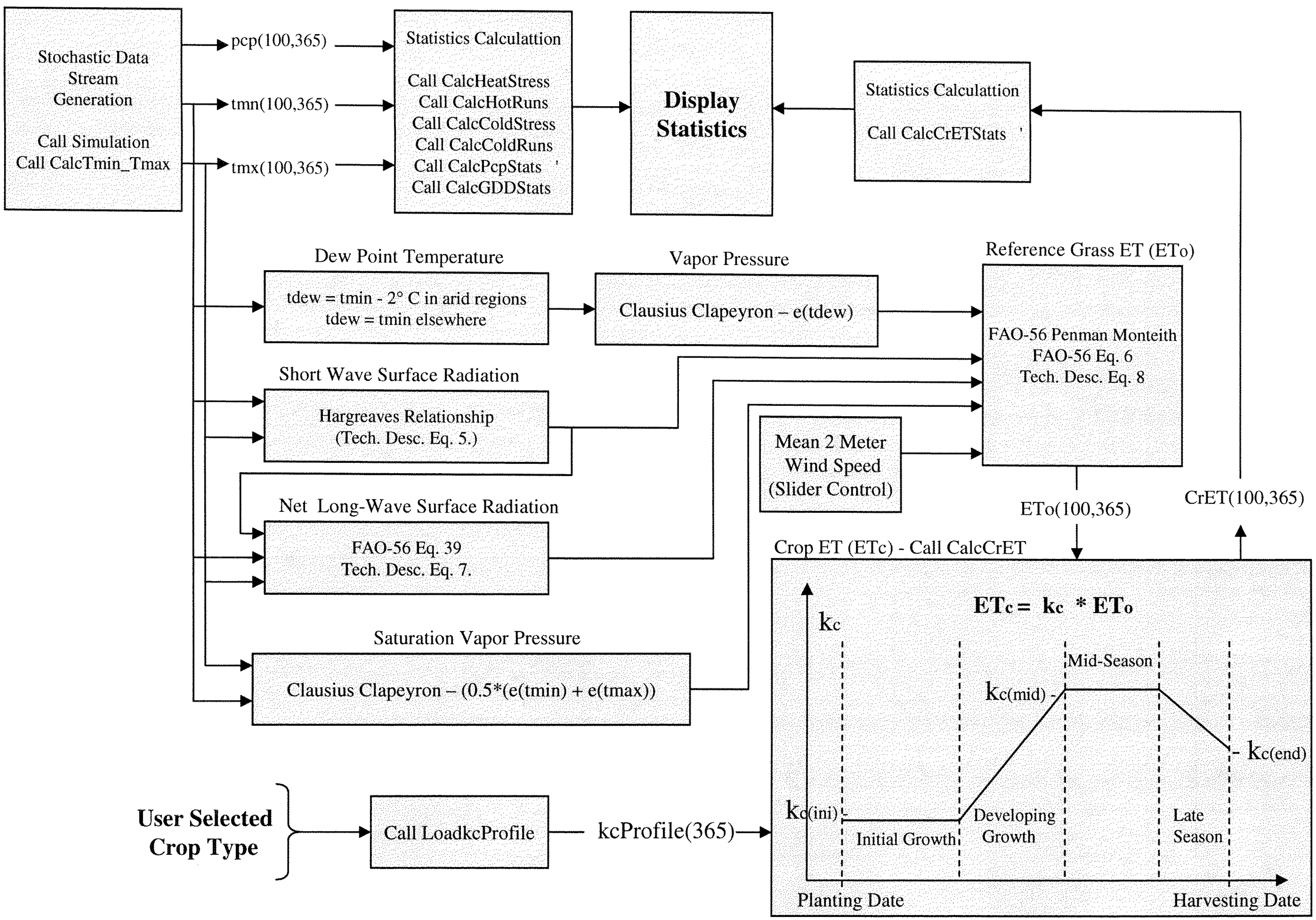
Net upwelling surface OLR was estimated using FAO-56 Eq. 39:

$$\mathbf{R_{nl}} = \sigma \left[\frac{\mathbf{T^4_{max}} + \mathbf{T^4_{min}}}{2} \right] \left(0.34 - 0.14\sqrt{\mathbf{e_a}}\right) \left(1.35 \frac{\mathbf{R_S}}{\mathbf{R_{SO}}} - 0.35\right) \tag{7}$$

Where,

- σ is the Stefan-Boltzmann constant,
- T_{min} is daily minimum temperature in Kelvin,

Fig. 6. ICARDA Agro Climate Tool Application Flow Chart: Secondary Synthetic Data Generation



- T_{max} is daily maximum temperature in Kelvin,
- R_S is the estimated daily shortwave radiation at the surface (Eq. 5)
- R_{SO} is the estimated clear sky daily shortwave radiation at the surface (FAO-56 Eq. 37), and,
- e_a is the actual vapor pressure (Eq. 6).

7.4 Reference Grass evapotranspiration

The FAO-56 method for deriving evapotranspiration rates for various crops is based on the estimation of reference evapotranspiration rates over a hypothetical grass surface (FAO-56, Chapter 4). Daily reference grass ET rates are calculated using the FAO-56 Penman-Monteith equation (FAO-56, Eq. 6).

$$\mathbf{ET_o} = \frac{0.408\mathbf{\Delta}\left((1-\alpha)\mathbf{R_S} - \mathbf{R_{nl}} - \mathbf{G}\right) + \boldsymbol{\gamma}\frac{900}{\mathbf{T}}\mathbf{u_2}(\mathbf{e_S} - \mathbf{e_a})}{\Delta + \gamma(1 + 0.34\mathbf{u_2})} \tag{8}$$

Where,

- Δ is the slope of the saturation vapor pressure at the mean daily temperature (FAO-56 Eq. 3-3),
- α is the albedo of the hypothetical grass surface (=.23),
- R_S is the shortwave solar radiation at the Earth's surface (Eq.),
- R_{nl} is the net upwelling outgoing long-wave radiation (OLR) at the surface (Eq.),
- G is the soil heat flux density (FAO-56 Eq. 5-2, with an assumed Leaf Area Index of 2.8),
- γ is the Psychometric constant (FAO-56 Eq. 8),
- T is the daily mean (i.e., $0.5(T_{max} + T_{min})$)temperature in Kelvin,
- u_2 is the mean wind speed at 2 meters, set her via a slider control on the Crop ET tab,
- e_a is the saturation vapor pressure, and,
- e_a is the actual vapor pressure.

7.5 Crop evapotranspiration

The application derives crop evapotranspiration rates over arbitrarily defined periods for a number of crops listed in the selection box at the top of the 'Crop ET' Tab. These crop ET rates (ET_c) are derived from a location's derived reference grass ET rates (ET_o) using the FAO-56 single crop coefficient method (FAO-56, Eq. 58).

$$\mathbf{ET_C} = \mathbf{k_C}\ \mathbf{ET_o} \tag{9}$$

Over the growing season crop ET is derived from the reference grass rates using k_c values drawn from a growing season coefficient profile (Fig. 1b). That coefficient profile is in turn derived from three k_c values defined during an initial crop growth period, a mid-season period, and an end of season value.

APPENDIX A: Statistical Sampling Error Assuming a Sample of 60 Measurements

The GEM6 weather generator parameters stored in columns 6-54 of the 'par_gen' table are the means, and Fourier amplitudes and phase angles of a meteorological station's annual cycles of temperature and rainfall variation. Those annual cycles are in turn formed from 12 statistics calculated over monthly periods:

- the means of daily minimum and maximum temperature (TMNM, TMXM),
- the standard deviation of minimum and maximum daily temperature (TMNS,TMXS),
- the average of daily rainfall on wet days (XMU), and,
- daily rainfall transition probabilities (p00, p10).

Because these statistics are calculated here from a sample of daily data containing as few as 60 measurements, they are subject to sampling error. In general, the magnitude of this sampling error is proportional to N^{-2}, where N is the number of measurements (Mendenhall et al., 1990).

For a sample mean calculated from 60 measurements (e.g., TMNM, TMXM, XMU), with a population mean of μ_x and a sample mean of x_s, sample means are normally distributed about the population mean.

$$\mu_x - x_s = \pm Z * \frac{\sigma_x}{\sqrt{60}} = \pm Z * .129 * \sigma_x \quad , \qquad \text{A.1}$$

where σ_x is the population standard deviation. As a result, with a ~68% probability (i.e., Z=1), the sampling error is less than or equal to .13 * σ_x. But in ~32% of cases, the error may be as large as .26 * σ_x (i.e., Z = 1.96). A probability estimate (P_s) derived from 60 measurements is similarly distributed about the true population probability value (P). The error between the true value and the estimate is given by,

$$P - P_s = \pm Z * \frac{\sqrt{P * (1 - P)}}{\sqrt{60}} \qquad \text{A.2}$$

For P = .5,

$$P - P_s = \pm Z * .5 * .129 = \pm Z * .065 \qquad \text{A.3}$$

Thus with a 68% probability, the sample probability mean will fall between .435 and .565. The error distribution narrows for values of P less than or greater than .5. For example, for P = .2 or .8,

$$P - P_s = \pm Z * .4 * .129 = \pm Z * .052 \qquad \text{A.4}$$

In that case the sample probability will fall within .148-.252 or .748-.852 with a 68% probability.

For a sample variance (S^2) calculated with 60 measurements and a population variance of σ^2, the ratio,

$$\frac{(\mathbf{n}-1)*\mathbf{S}^2}{\sigma^2}, \qquad \text{A.5}$$

is a χ^2 variable with 59 degrees of freedom (Mendenhall et al., (1990)) As a result, the ratio S^2/σ^2 lies within a specified range with a 95% certainty .

$$\mathbf{P}\left[\frac{43.188}{59} \leq \frac{\mathbf{S}^2}{\sigma^2} \leq \frac{79.082}{59}\right] = .95 \qquad \text{A.6}$$

Taking the square root,

$$\mathbf{P}\left[.732 \leq \frac{\mathbf{S}}{\sigma} \leq 1.15\right] = .95 \qquad \text{A.7}$$

So to within a 95% certainty, a sample standard deviation (S) calculated with 60 measurements lies within the range $.732*\sigma$ to $1.15*\sigma$, where σ is the true standard deviation.

Columns 55-62 in the Access table 'par_gen' are array elements which are in turn derived from correlation and cross-correlation values between daily minimum and maximum temperature. Like the sample mean and sample probability errors in Eqs. A.1 and A.2, sampling error for correlations estimated from a sample of 60 measurements is governed by an N^{-2} relationship (Box and Jenkins, 1976).

$$\rho - \rho_s = \pm\ \mathbf{Z}*\frac{1}{\sqrt{60}} = \pm\ \mathbf{Z}*.129 \qquad \text{A.8}$$

But unlike the sample probability relationship in Eq. A.2, the range of error does not narrow as the true correlation value (ρ) approaches 0.0 or 1.0. That is, if $\rho = .5$ then ρ_s will range between .371 and .629 with 68% probability. If $\rho = .2$ then ρ_s will still range +/- .129 of the true correlation value (.071-.329) with 68% probability. The relative error (i.e., $(\rho - \rho_s)/\rho$) is greater in the latter case.

8. Third Party Software Components

Continued development of the application's Visual Basic project ('beta4.vbp' under the source subdirectory on the source code CD ROM) will require acquiring the following software packages, and adding those packages on the project's Component Toolbox:

- ProEssentials Version 5 (Standard): Graphics software necessary to produce the application's various graph objects. Available at www.gigasoft.com.
- Component 6 Toolbox. A set of Active X controls that provides improved appearance relative to the controls normally provided in Microsoft Visual Studio. Available at:
  http://www.dbitech.com/product_page_Component_Toolbox.asp
- ESRI MapObjects Lite Version 2: Mapping and GIS software necessary to support the application's two maps. Available at:
  http://www.esri.com/software/mapobjectslt/index.html

Bibliography

Allen, R.G., Pereira, L.S., Raes, D., and Smith, M. 1998. *Crop evapotranspiration: Guidelines for computing crop water requirements.* FAO Irrig. and Drain. Paper No. 56, Food and Agriculture Organization of the United Nations, Rome, Italy. 300 p. (http://www.fao.org/docrep/X0490E/x0490e00.htm)

Box, G.P. and Jenkins, G.M. 1978. *Time series analysis: forecasting and control.* Holden-Day, Inc. Oakland, 818 p.

Hanson, C.L., Cummings, .A., Woolhiser, D.A., and Richardson, C.W. 1994. *Microcomputer program for daily weather simulation in the contiguous United States.* U.S. Department of Agriculture, Agricultural Research Service, ARS-114.

Mendenhall, W., Wackerly, D.D., and Scheaffer, R.L. 1990. *Mathematical Statistics with Applcationl.* PWS-Kent, Boston, 818 p.

National Climatic Data Center. 1994. Global Daily Summary: temperature and precipitation, 1977-1991. CD-ROM Version 1.0, Asheville, North Carolina.

National Climatic Data Center. 2002. The Global Daily Climatology Network (GDCN). CD-ROM Version 1.0, Asheville, North Carolina. (http://www.ncdc.noaa.gov/oa/climate/research/gdcn/gdcn.html).

Wilks, D.S. 1999. Interannual variability and extreme-value characteristics of several stochastic daily precipitation models. *Agricultural and Forest Meteorology*, 93: 153-169.